

Adaptive nonparametric estimation in heteroscedastic regression models.

Part 1: Sharp non-asymptotic oracle inequalities. *

L. Galtchouk [†] S. Pergamenshchikov[‡]

Abstract

An adaptive nonparametric estimation procedure is constructed for heteroscedastic regression when the noise variance depends on the unknown regression. A non-asymptotic upper bound for a quadratic risk (oracle inequality) is obtained.

Keywords: Adaptive estimation; Heteroscedastic regression; Nonasymptotic estimation; Nonparametric estimation; Oracle inequality.

AMS 2000 Subject Classifications: Primary: 62G08; Secondary: 62G05, 62G20

*The second author is partially supported by the RFFI-Grant 04-01-00855.

[†]Department of Mathematics, Strasbourg University 7, rue Rene Descartes, 67084, Strasbourg, France, e-mail: galtchou@math.u-strasbg.fr

[‡]Laboratoire de Mathématiques Raphael Salem, Avenue de l'Université, BP. 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex France, e-mail: Serge.Pergamenchikov@univ-rouen.fr

1 Introduction

Suppose we are given observations $(y_j)_{1 \leq j \leq n}$ which obey the heteroscedastic regression equation

$$y_j = S(x_j) + \sigma_j(S)\xi_j, \quad (1.1)$$

where design points $x_j = j/n$, $S(\cdot)$ is an unknown function to be estimated, $(\xi_j)_{1 \leq j \leq n}$ is a sequence of i.i.d. random variables, $(\sigma_j(S))_{1 \leq j \leq n}$ are unknown volatility coefficients depending on unknown regression function S .

The models of type (1.1) with $\sigma_j(S) = \sigma_j(x_j)$ were introduced in Akritas, Van Keilegom (2001) as a generalisation of the nonparametric ANCOVA model of Young and Bowman (1995). It should be noted that heteroscedastic regressions with this type of volatility coefficients have been encountered in econometric studies, namely, in consumer budget studies utilizing observations on individuals with diverse incomes and in analyses of the investment behavior of firms of different sizes (see Goldfeld, Quandt, 1972). For example, for consumer budget problems one uses there (see p. 83) some parametric version of model (1.1) with the volatility coefficient defined as

$$\sigma_j^2(S) = c_0 + c_1 x_j + c_2 S^2(x_j), \quad (1.2)$$

where c_0 , c_1 and c_2 are some positive unknown constants.

Moreover, this regression model appears in the drift estimation problem for stochastic differential equations when one passes from continuous time to discrete time model by making use of sequential kernel estimators having asymptotically minimal variances (see Galtchouk, Pergamenschikov, 2004; 2006; 2007a; 2007b).

The volatility coefficient estimation in heteroscedastic regression was considered in a few papers (see, for example, Cai, Wang, 2008 and the references therein). By making use of the squared first-order differences of the observations the initial problem in that paper was reduced to the regression function estimation in the model of type (1.1).

In this paper we develop the approach proposed in Galtchouk, Pergamenschikov (2005). The first goal of the research is to construct an adaptive procedure based on observations $(y_j)_{1 \leq j \leq n}$ for estimating the function S and to obtain a sharp non-asymptotic upper bound (oracle inequality) for a quadratic risk in the case when the smoothness of S is unknown. The second goal is to prove that the constructed procedure is efficient also in the asymptotic setup.

Problems of constructing a nonparametric estimator and proving a non-asymptotic upper bound for a risk in homoscedastic model, that is when $\sigma_j(S) \equiv \sigma$, were studied in few papers. A non-asymptotic upper bound for a quadratic risk over thresholding estimators is given in Kalifa, Mallat (2003). In papers by Barron, Birgé, Massart (1999), Massart (2004) an adaptive model selection procedure has been constructed. It is based on least squares estimators and a non-asymptotic upper bound has been obtained for a quadratic risk which is best in the principal term for the given class of estimators when the noise vector (ξ_1, \dots, ξ_n) is gaussian. This type of upper bounds is called the *oracle inequality*. In Fourdrinier, Pergamenschikov (2007) the oracle inequality has been obtained for a model selection procedure based on any estimators in the case when the noise vector (ξ_1, \dots, ξ_n) has a spherically symmetric distribution. Moreover, some sharp oracle inequalities have been obtained also for homoscedastic regression with gaussian noises, see, for example, Kneip (1994). Here the adjective "sharp" means that the coefficient of the principal term may be chosen as close to unity as desired.

In the paper for heteroscedastic regression an adaptive procedure is constructed for which the sharp non-asymptotic oracle inequality is proved. It should be noted that the methods used in former papers to obtain the sharp oracle inequality in regression models are limited by the homoscedastic case since they are based on the fact that an orthogonal transformation of a noise gaussian vector (ξ_1, \dots, ξ_n) gives a gaussian vector. In heteroscedastic regression models under consideration these methods are not valid since the noise vector is not gaussian. To obtain sharp non-asymptotic oracle inequalities in the heteroscedastic case the authors develop a new mathematical tools based on "penalty" methods and Pinsker's type weights.

Moreover, in Galtchouk, Pergamenschikov (2007c) we show that the given adaptive estimator is efficient in the asymptotic sense, that is, the sharp asymptotic lower bound is proved for a quadratic risk and it is attained over this estimator.

The paper is organized as follows. In Section 2 we construct an adaptive estimation procedure based on weighted least squares estimators and we obtain a non-asymptotic upper bound for the quadratic risk. In Section 3 we propose an estimator for the summarized noise variance and give the oracle inequality in the case of Sobolev space, $S \in W_r^k$. The proofs are given in Section 4. The Appendix contains some technical results.

2 Oracle inequality

In this paper we study the non-asymptotic estimation problem of the function S in the model (1.1) by observations $(y_j)_{1 \leq j \leq n}$ with odd sample number n . We assume that in (1.1) the sequence $(\xi_j)_{1 \leq j \leq n}$ is i.i.d. with

$$\mathbf{E}\xi_1 = 0, \quad \mathbf{E}\xi_1^2 = 1 \quad \text{and} \quad \mathbf{E}\xi_1^4 = \xi^* < \infty. \quad (2.1)$$

In the sequel we denote by $\bar{\xi} = \sqrt{\xi^* - 1}$.

Moreover, we assume that $(\sigma_l(S))_{1 \leq l \leq n}$ is a sequence of positive random variables independent of $(\xi_i)_{1 \leq i \leq n}$ and bounded away from $+\infty$, i.e. there exists some nonrandom unknown constant $\sigma_* \geq 1$ such that

$$\max_{1 \leq l \leq n} \sigma_l^2(S) \leq \sigma_*. \quad (2.2)$$

For any estimate \hat{S}_n of S based on observations $(y_j)_{1 \leq j \leq n}$, the estimation accuracy is measured by the mean integrated squared error (MISE)

$$\mathbf{E}_S \|\hat{S}_n - S\|_n^2, \quad (2.3)$$

where

$$\|\hat{S}_n - S\|_n^2 = (\hat{S}_n - S, \hat{S}_n - S)_n = \frac{1}{n} \sum_{l=1}^n (\hat{S}_n(x_l) - S(x_l))^2.$$

We make use of the trigonometric basis $(\phi_j)_{j \geq 1}$ in $\mathcal{L}_2[0, 1]$ with

$$\phi_1 = 1, \quad \phi_j(x) = \sqrt{2} \operatorname{Tr}_j(2\pi[j/2]x), \quad j \geq 2, \quad (2.4)$$

where the function $\operatorname{Tr}_j(x) = \cos(x)$ for even j and $\operatorname{Tr}_j(x) = \sin(x)$ for odd j ; $[x]$ denotes the integer part of x . Note that if n is odd, then this basis is orthonormal for the empirical inner product generated by the sieve $(x_j)_{1 \leq j \leq n}$, that is for any $1 \leq i, j \leq n$,

$$(\phi_i, \phi_j)_n = \frac{1}{n} \sum_{l=1}^n \phi_i(x_l) \phi_j(x_l) = \mathbf{K} \mathbf{r}_{ij}, \quad (2.5)$$

where $\mathbf{K} \mathbf{r}_{ij}$ is Kronecker's symbol.

By making use of this basis we define the discrete Fourier transformation in (1.1) and obtain the Fourier coefficients

$$\widehat{\theta}_{j,n} = (Y, \phi_j)_n \quad \text{and} \quad \theta_{j,n} = (S, \phi_j)_n. \quad (2.6)$$

Here $Y = (y_1, \dots, y_n)'$ and $S = (S(x_1), \dots, S(x_n))'$. The prime denotes the transposition.

From (1.1) it follows directly that these Fourier coefficients satisfy the following equation

$$\widehat{\theta}_{j,n} = \theta_{j,n} + \frac{1}{\sqrt{n}} \xi_{j,n} \quad (2.7)$$

with

$$\xi_{j,n} = \frac{1}{\sqrt{n}} \sum_{l=1}^n \sigma_l(S) \xi_l \phi_j(x_l).$$

We estimate the function S by the weighted least squares estimator

$$\widehat{S}_\lambda(x) = \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \phi_j(x), \quad (2.8)$$

where $x \in [0, 1]$, the weight vector $\lambda = (\lambda(1), \dots, \lambda(n))'$ belongs to some finite set Λ from $[0, 1]^n$. We denote by ν the cardinal number of the set Λ . Moreover, we set

$$\varrho_n = \max_{\lambda \in \Lambda} \sum_{j=1}^n \lambda(j) \quad \text{and} \quad \overline{\varrho}_{i,n} = \max_{\lambda \in \Lambda} \sup_{0 \leq x \leq 1} \left| \sum_{j=1}^n \lambda^i(j) \overline{\phi}_j(x) \right|, \quad (2.9)$$

where $\overline{\phi}_j = \phi_j^2 - 1$ and $i = 1, 2$.

Now we need to write a cost function to choose a weight $\lambda \in \Lambda$. Of course, it is obvious, that the best way is to minimize the cost function which is equal to the empirical squared error

$$\text{Err}_n(\lambda) = \|\widehat{S}_\lambda - S\|_n^2,$$

which in our case is equal to

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \widehat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \theta_{j,n} + \sum_{j=1}^n \theta_{j,n}^2. \quad (2.10)$$

Since coefficients $\theta_{j,n}$ are unknown, we need to replace the term $\widehat{\theta}_{j,n} \theta_{j,n}$ by some estimator which we choose as

$$\widetilde{\theta}_{j,n} = \widehat{\theta}_{j,n}^2 - \frac{1}{n} \widehat{\varsigma}_n,$$

where $\widehat{\varsigma}_n$ is some estimator of the summarized noise variance

$$\varsigma_n = n^{-1} \sum_{l=1}^n \sigma_l^2(S). \quad (2.11)$$

Such type of estimators is given in (3.5).

Moreover, for this substitution to the empirical squared error one needs to pay a penalty. Finally, we define the cost function by the following way

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \widehat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \widetilde{\theta}_{j,n} + \rho \widehat{P}_n(\lambda), \quad (2.12)$$

where ρ is some positive coefficient which will be chosen later. The penalty term we define as

$$\widehat{P}_n(\lambda) = \frac{|\lambda|^2 \widehat{\varsigma}_n}{n} \quad \text{with} \quad |\lambda|^2 = \sum_{j=1}^n \lambda^2(j). \quad (2.13)$$

Note that in the case when the sequence $(\sigma_l(S))_{1 \leq l \leq n}$ is known, i.e. $\widehat{\varsigma}_n = \varsigma_n$, we obtain

$$P_n(\lambda) = \frac{|\lambda|^2 \varsigma_n}{n}. \quad (2.14)$$

We set

$$\widehat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_n(\lambda) \quad (2.15)$$

and define an estimator of S as

$$\widehat{S}_* = \widehat{S}_{\widehat{\lambda}}. \quad (2.16)$$

We recall that the set Λ is finite so $\widehat{\lambda}$ exists. In the case when $\widehat{\lambda}$ is not unique we take one of them.

To formulate the oracle inequality we introduce, for $0 < \rho < 1/3$, the following function

$$\Psi_n(\rho) = \frac{\rho(1-\rho)\Upsilon_n^*(\rho) + 2\nu + 2\rho^2(1-\rho)\overline{\varrho}_{2,n}}{\rho(1-3\rho)} \sigma_* \quad (2.17)$$

with

$$\Upsilon_n^*(\rho) = \frac{16\nu}{\rho} + 4\bar{\varrho}_{1,n} \left(1 + \nu \frac{\bar{\xi}}{\sqrt{n}} \right) + 4\nu \varrho_n \frac{\bar{\xi}}{\sqrt{n}}.$$

Theorem 2.1. *Let Λ be any finite set in $[0, 1]^n$. For any $n \geq 3$ and $0 < \rho < 1/3$, the estimator \hat{S}_* satisfies the oracle inequality*

$$\mathbf{E}_S \|\hat{S}_* - S\|_n^2 \leq (1 + \kappa(\rho)) \min_{\lambda \in \Lambda} \mathbf{E}_S \|\hat{S}_\lambda - S\|_n^2 + \frac{1}{n} \mathcal{B}_n(\rho), \quad (2.18)$$

where $\mathcal{B}_n(\rho) = \Psi_n(\rho) + \kappa_*(\rho) \varrho_n \mathbf{E}_S |\hat{\varsigma}_n - \varsigma_n|$ with

$$\kappa(\rho) = \frac{6\rho - 2\rho^2}{1 - 3\rho} \quad \text{and} \quad \kappa_*(\rho) = 4 \frac{1 - \rho^2}{1 - 3\rho}.$$

If in model (1.1) the volatility coefficients $(\sigma_l(S))_{1 \leq l \leq n}$ are known, then $\hat{\varsigma}_n = \varsigma_n$ and inequality (2.18) has the following form

$$\mathbf{E}_S \|\hat{S}_* - S\|_n^2 \leq (1 + \kappa(\rho)) \min_{\lambda \in \Lambda} \mathbf{E}_S \|\hat{S}_\lambda - S\|_n^2 + \frac{1}{n} \Psi_n(\rho). \quad (2.19)$$

Remark 2.1. *Note that the principal term in the right-hand side of (2.18)-(2.19) is best in the class of estimators $(\hat{S}_\lambda, \lambda \in \Lambda)$. Inequalities of such type are called the sharp non-asymptotic oracle inequalities. The inequality is sharp in the sense that the coefficient of the principal term may be chosen as close to 1 as desired. Similar inequalities for homoscedastic models (1.1) with $\sigma_l(S) = \sigma$ were given, for example, in [13]. The methods used there cannot be extended to the heteroscedastic case since, after the Fourier transformation, the random variables $(\xi_{i,n})$ in model (2.7) are dependent contrary to the homoscedastic case, where these random variables are independent (see, for example, Rohde, 2004).*

Remark 2.2. *If one would like to obtain the asymptotically minimal MISE of the estimator \hat{S}_* , then the secondary term $\mathcal{B}_n(\rho)$ in (2.18) should be slowly varying. Indeed, since usually the optimal rate is of order $n^{2k/(2k+1)}$ for some $k \geq 1$, then after multiplying the inequality (2.18) by this rate the principal term gives the optimal constant and the secondary one should be of type that for any $\delta > 0$*

$$\frac{\mathcal{B}_n(\rho)}{n^\delta} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

Due to the definitions $\Psi_n(\rho)$ and $\mathcal{B}_n(\rho)$, it should be, for any $\delta > 0$,

$$\rho n^\delta \rightarrow +\infty, \quad \frac{\varrho_n \mathbf{E}_S |\widehat{\varsigma}_n - \varsigma_n|}{n^\delta} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

One can take, for example, the parameter ρ tending to zero as $n \rightarrow \infty$ like

$$\rho = O\left(\frac{1}{\ln^\gamma n}\right) \quad (2.20)$$

for some $\gamma > 0$. The choice of ϱ_n and of the estimator $\widehat{\varsigma}_n$ is proposed below.

Consider now the order of the termes $\varrho_n, \overline{\varrho}_{1,n}, \overline{\varrho}_{2,n}$ and the function $\Psi_n(\rho)$ in the case when the finite set Λ is formed by a special version of Pinsker's weights (see, for example, [15]). To this end, we define the sieve

$$\mathcal{A}_\varepsilon = \{1, \dots, k_*\} \times \{t_1, \dots, t_m\},$$

where $t_i = i\varepsilon$ and $m = \lceil 1/\varepsilon^2 \rceil$. We suppose that the parameters $k_* \geq 1$ and $0 < \varepsilon \leq 1$ are functions of n such that,

$$\begin{cases} \lim_{n \rightarrow \infty} k_* = +\infty, & \lim_{n \rightarrow \infty} \frac{k_*}{\ln n} = 0, \\ \lim_{n \rightarrow \infty} \varepsilon = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^\delta \varepsilon = +\infty, \end{cases} \quad (2.21)$$

for any $\delta > 0$. For example, one can take $\varepsilon = 1/\ln n$ and $k_* = \sqrt{\ln n}$ for $n \geq 3$.

For any $\alpha = (\beta, t) \in \mathcal{A}_\varepsilon$ we define the weight vector $\lambda_\alpha = (\lambda_\alpha(1), \dots, \lambda_\alpha(n))'$ as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j \leq j_0\}} + (1 - (j/\omega_\alpha)^\beta) \mathbf{1}_{\{j_0 < j \leq \omega_\alpha\}}, \quad (2.22)$$

where $j_0 = j_0(\alpha) = \lfloor \omega_\alpha / \ln n \rfloor$,

$$\omega_\alpha = (A_\beta t n)^{1/(2\beta+1)} \quad \text{and} \quad A_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}.$$

Hence,

$$\Lambda = \{\lambda_\alpha, \alpha \in \mathcal{A}_\varepsilon\} \quad (2.23)$$

and $\nu = k_* m$. Note that in this case in view of (2.21) for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\nu}{n^\delta} = 0.$$

Moreover, by (2.22)

$$\sum_{j=1}^n \lambda_\alpha(j) = \mathbf{1}_{\{j_0 \geq 1\}} j_0 + \mathbf{1}_{\{\omega_\alpha \geq 1\}} \sum_{j=j_0+1}^{[\omega_\alpha]} (1 - (j/\omega_\alpha)^\beta) \leq \omega_\alpha.$$

Therefore, taking into account that $A_\beta \leq A_1 < 1$ for $\beta \geq 1$ we find that

$$\varrho_n \leq (n/\varepsilon)^{1/3},$$

i.e. for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\varrho_n}{n^{1/3+\delta}} = 0.$$

Moreover, note that for any $x \in [0, 1]$, we get

$$\begin{aligned} \sum_{j=1}^n \lambda_\alpha(j) \bar{\phi}_j(x) &= \mathbf{1}_{\{j_0 \geq 1\}} \sum_{j=1}^{j_0} \bar{\phi}_j(x) \\ &\quad + \mathbf{1}_{\{\omega_\alpha \geq 1\}} \sum_{j=j_0+1}^{[\omega_\alpha]} (1 - (j/\omega_\alpha)^\beta) \bar{\phi}_j(x). \end{aligned}$$

Thus Lemma A.2 implies that

$$\bar{\varrho}_{1,n} \leq 1 + 2^{\beta+1} \leq 1 + 2^{k_*+1}.$$

Due to the condition for k_* in (2.21) this function is slowly varying, i.e. for any $\delta > 0$,

$$\lim_{n \rightarrow \infty} \frac{\bar{\varrho}_{1,n}}{n^\delta} = 0.$$

By the same way we obtain that

$$\bar{\varrho}_{2,n} \leq 1 + 2^{k_*+2} + 2^{2k_*+1}$$

and, therefore, for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\bar{\varrho}_{2,n}}{n^\delta} = 0.$$

Thus, if we choose the parameter $\rho = \rho_n$ as in (2.20) we obtain that in this case, for any $\delta > 0$,

$$\lim_{n \rightarrow \infty} \frac{\Psi_n(\rho)}{n^\delta} = 0. \tag{2.24}$$

3 Oracle inequality for $S \in W_r^k$

Assume that $S : \mathbb{R} \rightarrow \mathbb{R}$ is a k times differentiable 1-periodic function such that

$$\sum_{j=0}^k \|S^{(j)}\|^2 \leq r, \quad (3.1)$$

where

$$\|f\|^2 = \int_0^1 f^2(t) dt. \quad (3.2)$$

We denote by W_r^k the set of all such functions. Moreover, we suppose that $r > 0$ and $k \geq 1$ are unknown parameters.

Note that, the space W_r^k can be represented as an ellipses in the Hilbert space, i.e.

$$W_r^k = \{S \in \mathcal{L}_2[0, 1] : S = \sum_{j=1}^{\infty} \theta_j \phi_j \text{ such that } \sum_{j=1}^{\infty} a_j \theta_j^2 \leq r\}, \quad (3.3)$$

where the basis functions $(\phi_j)_{j \geq 1}$ are defined in (2.4); $(\theta_j)_{j \geq 1}$ are the Fourier coefficients, i.e.

$$\theta_j = (S, \phi_j) = \int_0^1 S(t) \phi_j(t) dt. \quad (3.4)$$

The coefficients $(a_j)_{j \geq 1}$ are defined as

$$a_j = \sum_{l=0}^k \|\phi_j^{(l)}\|^2 = \sum_{l=0}^k (2\pi[j/2])^{2l}.$$

To estimate ς_n , we make use of the following estimator:

$$\widehat{\varsigma}_n = \sum_{j=m_n+1}^n \widehat{\theta}_{j,n}^2, \quad (3.5)$$

where the parameter $1 \leq m_n \leq n$ will be chosen later.

In Section 4 we show the following result.

Lemma 3.1. *For any $n \geq 2$ and $r > 0$,*

$$\sup_{S \in W_r^1} \mathbf{E}_S |\widehat{\varsigma}_n - \varsigma_n| \leq \frac{\overline{\sigma} + \varsigma_n^*(r)}{\sqrt{n}}, \quad (3.6)$$

where $\bar{\sigma} = 2(\bar{\xi} + \sqrt{2})\sigma_*$ and

$$\varsigma_n^*(r) = \frac{4r\sqrt{n}}{m_n^2} + 4\sqrt{r\sigma_*}\frac{1}{m_n} + \frac{(2 + m_n)\sigma_*}{\sqrt{n}}.$$

If we choose the parameter m_n in (3.5) such that

$$\lim_{n \rightarrow \infty} \frac{m_n}{\sqrt{n}} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{m_n^2}{\sqrt{n}} = \infty, \quad (3.7)$$

we obtain that

$$\lim_{n \rightarrow \infty} \varsigma_n^*(r) = 0.$$

Theorem 2.1 and inequality (3.6) imply immediately the following result.

Theorem 3.2. *Let Λ be any finite set in $[0, 1]^n$. Assume that in the model (1.1) the function S belongs to W_r^1 . Then, for any $n \geq 3$ and $0 < \rho < 1/3$, the procedure \hat{S}_* from (2.16) with $\hat{\varsigma}_n$ defined by (3.5) and (3.7) satisfies the following oracle inequality*

$$\mathbf{E}_S \|\hat{S}_* - S\|_n^2 \leq (1 + \kappa(\rho)) \min_{\lambda \in \Lambda} \mathbf{E}_S \|\hat{S}_\lambda - S\|_n^2 + \frac{1}{n} \mathcal{D}_n(\rho, r), \quad (3.8)$$

where

$$\mathcal{D}_n(\rho, r) = \Psi_n(\rho) + \kappa_*(\rho) (\bar{\sigma} + \varsigma_n^*(r)) \frac{\varrho_n}{n}.$$

If the set Λ is from (2.23), then for any $\delta > 0$ and any $0 < \rho < 1/3$

$$\lim_{n \rightarrow \infty} \frac{\mathcal{D}_n(\rho, r)}{n^\delta} = 0.$$

4 Proofs

4.1 Proof of Theorem 2.1

First of all, note that we can represent the empirical squared error $\text{Err}_n(\lambda)$ by the following way

$$\text{Err}_n(\lambda) = J_n(\lambda) + 2 \sum_{j=1}^n \lambda(j) \theta'_{j,n} + \|S\|_n^2 - \rho \hat{P}_n(\lambda) \quad (4.1)$$

with $\theta'_{j,n} = \tilde{\theta}_{j,n} - \theta_{j,n} \hat{\theta}_{j,n}$. By setting

$$\varsigma_{j,n} = \mathbf{E}_S \xi_{j,n}^2 = \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S) \phi_j^2(x_l), \quad (4.2)$$

we find that

$$\theta'_{j,n} = \frac{1}{\sqrt{n}} \theta_{j,n} \xi_{j,n} + \frac{1}{n} \tilde{\xi}_{j,n} + \frac{1}{n} \tilde{\delta}_{j,n},$$

where

$$\tilde{\xi}_{j,n} = \xi_{j,n}^2 - \varsigma_{j,n} \quad \text{and} \quad \tilde{\delta}_{j,n} = \varsigma_{j,n} - \hat{\varsigma}_n. \quad (4.3)$$

Note now that, we can represent $\tilde{\xi}_{j,n}$ as

$$\tilde{\xi}_{j,n} = \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S) \phi_j^2(x_l) \tilde{\xi}_l + 2 \sum_{l=2}^n \tau_{j,l} \xi_l = \tilde{\xi}'_{j,n} + 2 \tilde{\xi}''_{j,n}, \quad (4.4)$$

where $\tilde{\xi}_l = \xi_l^2 - 1$ and

$$\tau_{j,l} = \frac{1}{n} \sigma_l(S) \phi_j(x_l) \sum_{d=1}^{l-1} \sigma_d(S) \phi_j(x_d) \xi_d.$$

Now we set

$$N_1(\lambda) = \sum_{j=1}^n \lambda(j) \tilde{\xi}'_{j,n} \quad \text{and} \quad N_2(\lambda) = \frac{1}{\sqrt{n \varsigma_n}} \sum_{j=1}^n \bar{\lambda}(j) \tilde{\xi}''_{j,n} \mathbf{1}_{\{\varsigma_n > 0\}}, \quad (4.5)$$

where $\bar{\lambda}(j) = \lambda(j)/|\lambda|$. In the Appendix we show that

$$\sup_{\lambda \in \Lambda} \mathbf{E}_S |N_1(\lambda)| \leq \bar{\xi} \sigma_*(\varrho_n + \bar{\varrho}_{1,n}) \frac{1}{\sqrt{n}} \quad (4.6)$$

and

$$\sup_{\lambda \in \mathbb{R}^n} \mathbf{E}_S (N_2(\lambda))^2 \leq 2\sigma_*/n. \quad (4.7)$$

Now, for any $\lambda \in \Lambda$, we rewrite (4.1) as

$$\begin{aligned} \text{Err}_n(\lambda) &= J_n(\lambda) + \frac{2}{n} N_1(\lambda) + 4\sqrt{P_n(\lambda)} N_2(\lambda) \\ &\quad + 2M(\lambda) + \frac{2}{n} \tilde{\Delta}(\lambda) + \|S\|_n^2 - \rho \hat{P}_n(\lambda), \end{aligned}$$

where $P_n(\lambda)$ is defined in (2.14),

$$\tilde{\Delta}(\lambda) = \sum_{j=1}^n \lambda(j) \tilde{\delta}_{j,n} \quad \text{and} \quad M(\lambda) = n^{-1/2} \sum_{j=1}^n \lambda(j) \theta_{j,n} \xi_{j,n}. \quad (4.8)$$

We start with $\tilde{\Delta}(\lambda)$. Setting

$$\bar{\varsigma}_{j,n} = \varsigma_{j,n} - \varsigma_n = \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S) \bar{\phi}_j(x_l), \quad (4.9)$$

we obtain that

$$\begin{aligned} |\tilde{\Delta}(\lambda)| &\leq \left| \sum_{j=1}^n \lambda(j) \bar{\varsigma}_{j,n} \right| + \varrho_n |\hat{\varsigma}_n - \varsigma_n| \\ &\leq \sigma_* \bar{\varrho}_{1,n} + \varrho_n |\hat{\varsigma}_n - \varsigma_n|. \end{aligned} \quad (4.10)$$

Now from (4.1) we obtain that, for some fixed $\lambda_0 \in \Lambda$,

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) - \text{Err}_n(\lambda_0) &= J(\hat{\lambda}) - J(\lambda_0) + 2M(\hat{\vartheta}) + \frac{2}{n} N_1(\hat{\vartheta}) \\ &\quad + 4\sqrt{P_n(\hat{\lambda})} N_2(\hat{\lambda}) - 4\sqrt{P_n(\lambda_0)} N_2(\lambda_0) \\ &\quad - \rho \hat{P}_n(\hat{\lambda}) + \rho \hat{P}_n(\lambda_0) + \frac{2}{n} \left(\tilde{\Delta}(\hat{\lambda}) - \tilde{\Delta}(\lambda_0) \right), \end{aligned}$$

where $\hat{\vartheta} = \hat{\lambda} - \lambda_0$.

By the definition of $\hat{\lambda}$ in (2.15) and by (4.10) we get

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) - \text{Err}_n(\lambda_0) &\leq 2M(\hat{\vartheta}) + \frac{4\sigma_* \bar{\varrho}_{1,n} + 4\varrho_n |\hat{\varsigma}_n - \varsigma_n|}{n} \\ &\quad + \frac{2}{n} N_1(\hat{\vartheta}) + 4\sqrt{P_n(\hat{\lambda})} N_2(\hat{\lambda}) - \rho \hat{P}_n(\hat{\lambda}) \\ &\quad + \rho \hat{P}_n(\lambda_0) - 4\sqrt{P_n(\lambda_0)} N_2(\lambda_0). \end{aligned}$$

Moreover, making use of the inequality

$$2|ab| \leq \varepsilon a^2 + \varepsilon^{-1} b^2 \quad (4.11)$$

with $\varepsilon = \rho/4$ and taking into account the definition of penalty term in (2.13) we deduce, for any $\lambda \in \Lambda$,

$$\begin{aligned} 4\sqrt{P_n(\lambda)}|N_2(\lambda)| &\leq \rho P_n(\lambda) + 4\frac{N_2^2(\lambda)}{\rho} \\ &\leq \rho\widehat{P}_n(\lambda) + \rho\frac{|\lambda|^2|\widehat{\varsigma}_n - \varsigma_n|}{n} + \frac{4N_2^2(\lambda)}{\rho}. \end{aligned}$$

Thus from here it follows that

$$\text{Err}_n(\widehat{\lambda}) \leq \text{Err}_n(\lambda_0) + 2M(\widehat{\vartheta}) + \Upsilon_n + 2\rho\widehat{P}_n(\lambda_0), \quad (4.12)$$

where

$$\Upsilon_n = \frac{4}{n}N_1^* + \frac{8}{\rho}(N_2^*)^2 + \frac{4\sigma_*\overline{\varrho}_{1,n}}{n} + \frac{4+2\rho}{n}\varrho_n|\widehat{\varsigma}_n - \varsigma_n|$$

with $N_1^* = \sup_{\lambda \in \Lambda} |N_1(\lambda)|$ and $N_2^* = \sup_{\lambda \in \Lambda} |N_2(\lambda)|$. Moreover, note that the bounds (4.6), (4.7) and (4.10) imply that

$$\mathbf{E}_S \Upsilon_n \leq \Upsilon_n^*(\rho) \frac{\sigma_*}{n} + \frac{4+2\rho}{n} \varrho_n \mathbf{E}_S |\widehat{\varsigma}_n - \varsigma_n|, \quad (4.13)$$

where the function $\Upsilon_n^*(\rho)$ is defined in (2.17).

Now we study the second term in (4.8). First, note that for any nonrandom vector $\vartheta = (\vartheta(1), \dots, \vartheta(n))' \in \mathbb{R}^n$ Lemma A.4 implies

$$\mathbf{E}_S M^2(\vartheta) \leq \frac{\sigma_*}{n} \sum_{j=1}^n \vartheta^2(j) \theta_{j,n}^2 = \sigma_* \frac{\|S_\vartheta\|_n^2}{n}, \quad (4.14)$$

where

$$S_\vartheta = \sum_{j=1}^n \vartheta(j) \theta_{j,n} \phi_j.$$

We set now

$$Z^* = \sup_{\vartheta \in \Lambda_1} \frac{nM^2(\vartheta)}{\|S_\vartheta\|_n^2} \quad \text{with} \quad \Lambda_1 = \Lambda - \lambda_0.$$

We estimate this term with the help of inequality (4.14), i.e.

$$\mathbf{E}_S Z^* \leq \sum_{\vartheta \in \Lambda_1} \frac{n\mathbf{E}_S M^2(\vartheta)}{\|S_\vartheta\|_n^2} \leq \nu\sigma_*. \quad (4.15)$$

Moreover, making use of inequality (4.11) with $\varepsilon = \rho \|S_\vartheta\|_n$, we get

$$2|M(\vartheta)| \leq \rho \|S_\vartheta\|_n^2 + \frac{Z^*}{n\rho}. \quad (4.16)$$

Now we estimate $\|S_\vartheta\|_n^2$. We have

$$\|S_\vartheta\|_n^2 - \|\widehat{S}_\vartheta\|_n^2 = \sum_{j=1}^n \vartheta^2(j) (\theta_{j,n}^2 - \widehat{\theta}_{j,n}^2) \leq -2M_1(\vartheta) \quad (4.17)$$

with

$$M_1(\vartheta) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \vartheta^2(j) \theta_{j,n} \xi_{j,n}.$$

Now, taking into account that $|\vartheta(j)| \leq 1$ for any $\vartheta \in \Lambda_1$, we obtain

$$\mathbf{E}_S M_1^2(\vartheta) \leq \sigma_* \frac{\|S_\vartheta\|_n^2}{n}.$$

Putting

$$Z_1^* = \sup_{\vartheta \in \Lambda_1} \frac{nM_1^2(\vartheta)}{\|S_\vartheta\|_n^2},$$

we get

$$\mathbf{E}_S Z_1^* \leq \nu \sigma_*. \quad (4.18)$$

Therefore, applying inequality (4.16) for $M_1(\vartheta)$ in (4.17) we deduce the upper bound for $\|S_\vartheta\|_n^2$, i.e.

$$\|S_\vartheta\|_n^2 \leq \frac{1}{1-\rho} \|\widehat{S}_\vartheta\|_n^2 + \frac{Z_1^*}{n\rho(1-\rho)}. \quad (4.19)$$

Taking into account this inequality in (4.16) we obtain that

$$\begin{aligned} 2M(\vartheta) &\leq \frac{\rho}{1-\rho} \|\widehat{S}_\vartheta\|_n^2 + \frac{Z^* + Z_1^*}{n\rho(1-\rho)} \\ &\leq \frac{2\rho(\text{Err}_n(\lambda) + \text{Err}_n(\lambda_0))}{1-\rho} + \frac{Z^* + Z_1^*}{n\rho(1-\rho)}. \end{aligned}$$

Therefore (4.12) implies that

$$\begin{aligned}\text{Err}_n(\widehat{\lambda}) &\leq \frac{1+\rho}{1-3\rho}\text{Err}_n(\lambda_0) + \frac{1-\rho}{1-3\rho}\Upsilon_n \\ &\quad + \frac{Z^* + Z_1^*}{n\rho(1-3\rho)} + \frac{2\rho(1-\rho)}{1-3\rho}\widehat{P}_n(\lambda_0),\end{aligned}$$

Now by inequalities (4.15)–(4.18) we get that

$$\begin{aligned}\mathbf{E}_S\text{Err}_n(\widehat{\lambda}) &\leq \frac{1+\rho}{1-3\rho}\mathbf{E}_S\text{Err}_n(\lambda_0) + \frac{1-\rho}{1-3\rho}\mathbf{E}_S\Upsilon_n \\ &\quad + \frac{2\nu\sigma_*}{n\rho(1-3\rho)} + \frac{2\rho(1-\rho)}{1-3\rho}\mathbf{E}_S\widehat{P}_n(\lambda_0).\end{aligned}$$

By making use of inequality (4.13) and Lemma A.1 we come to Theorem 2.1.

□

4.2 Proof of Lemma 3.1

First notice that from (2.7) we obtain that

$$\begin{aligned}\widehat{\varsigma}_n - \varsigma_n &= \sum_{j=m_n+1}^n \theta_{j,n}^2 + \frac{2}{\sqrt{n}} \sum_{j=m_n+1}^n \theta_{j,n} \xi_{j,n} \\ &\quad + n^{-1} \sum_{j=m_n+1}^n \widetilde{\xi}_{j,n} + n^{-1} \sum_{j=m_n+1}^n \overline{\varsigma}_{j,n} - \frac{m_n}{n} \varsigma_n \\ &:= \Delta_1 + \frac{2}{\sqrt{n}} \Delta_2 + \frac{1}{n} \Delta_3 + \frac{1}{n} \Delta_4 - \frac{m_n}{n} \varsigma_n,\end{aligned}$$

where $\widetilde{\xi}_{j,n}$ and $\overline{\varsigma}_{j,n}$ are defined in (4.3) and (4.9) respectively.

We estimate the first term by Lemma A.3 for $S \in W_r^1$. We have

$$\Delta_1 \leq \frac{4r}{m_n^2}.$$

The next term we estimate with the help of Lemma A.4. We get that

$$\mathbf{E}_S(\Delta_2)^2 \leq \sigma_* \Delta_1 \leq \sigma_* \frac{4r}{m_n^2}.$$

By (4.4) and (4.5) we can represent Δ_3 as

$$\Delta_3 = N_1(\lambda_I) + 2|\lambda_I|\sqrt{n\bar{\varsigma}_n}N_2(\lambda_I)$$

with the vector $\lambda_I = (\lambda_I(1), \dots, \lambda_I(n))'$ having the indicator components, i.e. $\lambda_I(j) = \mathbf{1}_{\{j > m_n\}}$. By estimating in (A.1) ϕ_j^2 by 2 we obtain

$$\mathbf{E}_S |N_1(\lambda_I)| \leq 2\sigma_* \bar{\xi} \sqrt{n}.$$

Thus the upper bound (4.7) implies

$$\mathbf{E}_S |\Delta_3| \leq 2\sigma_*(\bar{\xi} + \sqrt{2})\sqrt{n} = \bar{\sigma}\sqrt{n}.$$

Moreover, due to Lemma A.2 with $m = 0$, one has

$$\begin{aligned} |\Delta_4| &= \left| n^{-1} \sum_{d=1}^n \sigma_d^2(S) \sum_{j=m_n+1}^n \bar{\phi}_j(x_d) \right| \\ &\leq \frac{\sigma_*}{n} \sum_{d=1}^n \left| \sum_{j=1}^n \bar{\phi}_j(x_d) \right| + \frac{\sigma_*}{n} \sum_{d=1}^n \left| \sum_{j=1}^{m_n} \bar{\phi}_j(x_d) \right| \leq 2\sigma_*. \end{aligned}$$

Hence Lemma 3.1. \square

5 Appendix

A.1 Proof of (4.6)

First note that we can represent the term $N_1(\lambda)$ as

$$N_1(\lambda) = \sum_{l=1}^n v_{l,n} \tilde{\xi}_l \quad \text{with} \quad v_{l,n} = \frac{\sigma_l^2(S)}{n} \sum_{j=1}^n \lambda(j) \phi_j^2(x_l).$$

Recalling that $\mathbf{E} \tilde{\xi}_1^2 = \xi^* - 1 = \bar{\xi}^2$ we calculate

$$\mathbf{E}_S N_1^2(\lambda) = \frac{\bar{\xi}^2}{n^2} \sum_{l=1}^n \mathbf{E}_S \sigma_l^4(S) \left(\sum_{j=1}^n \lambda(j) \phi_j^2(x_l) \right)^2.$$

Therefore for any vector $\lambda \in \mathbb{R}^n$

$$\mathbf{E}_S |N_1(\lambda)| \leq \sigma_* \frac{\bar{\xi}}{\sqrt{n}} \max_{0 \leq x \leq 1} \left| \sum_{j=1}^n \lambda(j) \phi_j^2(x) \right|. \quad (\text{A.1})$$

Thus taking into account here definitions (2.9) we come to inequality (4.6).
 \square

A.2 Proof of (4.7)

By putting $\alpha_l = \sum_{j=1}^n \bar{\lambda}(j) \tau_{j,l}$ and taking into account that the random variables $(\xi_k)_{1 \leq k \leq n}$ are independent of $(\sigma_k(S))_{1 \leq k \leq n}$ we obtain that

$$\mathbf{E}_S (N_2^2(\lambda) | \sigma_k(S), 1 \leq k \leq n) = \mathbf{1}_{\{\varsigma_n > 0\}} \left(\sum_{l=1}^n \sigma_l^2(S) \right)^{-1} \sum_{j=1}^n \hat{\alpha}_l, \quad (\text{A.2})$$

where

$$\begin{aligned} \hat{\alpha}_l &= \mathbf{E}(\alpha_l^2 | \sigma_k(S), 1 \leq k \leq n) \\ &= \frac{\sigma_l^2(S)}{n^2} \sum_{r=1}^{l-1} \sigma_r^2(S) \left(\sum_{j=1}^n \bar{\lambda}(j) \phi_j(x_l) \phi_j(x_r) \right)^2. \end{aligned}$$

Therefore the orthonormality property (2.5) implies that for any $\lambda \in \mathbb{R}^n$

$$\begin{aligned} \hat{\alpha}_l &\leq \sigma_* \frac{\sigma_l^2(S)}{n^2} \sum_{d=1}^n \left(\sum_{j=1}^n \bar{\lambda}(j) \phi_j(x_l) \phi_j(x_d) \right)^2 \\ &= \sigma_* \frac{\sigma_l^2(S)}{n} \sum_{j=1}^n \bar{\lambda}^2(j) \phi_j^2(x_l) \leq \frac{2\sigma_*}{n} \sigma_l^2(S). \end{aligned}$$

Now by making use of this inequality in (A.2) we get (4.7).
 \square

A.3 Technical lemma

Lemma A.1. *For any $n \geq 1$ and $\lambda \in \Lambda$,*

$$\mathbf{E}_S \widehat{P}_n(\lambda) \leq \mathbf{E}_S \text{Err}_n(\lambda) + \frac{\varrho_n}{n} \mathbf{E}_S |\widehat{\varsigma}_n - \varsigma_n| + \frac{\sigma_* \bar{\varrho}_{2,n}}{n}.$$

Proof. Indeed, by the definition of $\text{Err}_n(\lambda)$ we have

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \left((\lambda(j) - 1) \theta_{j,n} + \lambda(j) \frac{1}{\sqrt{n}} \xi_{j,n} \right)^2.$$

Therefore,

$$\mathbf{E}_S \text{Err}_n(\lambda) \geq \mathbf{E}_S \frac{1}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 = \mathbf{E}_S \frac{1}{n} \sum_{j=1}^n \lambda^2(j) \varsigma_{j,n},$$

where the sequence $(\varsigma_{j,n})$ is defined in (4.2). Moreover, note that the last term can be estimated as

$$\left| \sum_{j=1}^n \lambda^2(j) \varsigma_{j,n} - |\lambda|^2 \varsigma_n \right| = \left| \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S) \sum_{j=1}^n \lambda^2(j) \bar{\phi}_j(x_l) \right| \leq \sigma_* \bar{\varrho}_{2,n}.$$

We recall that the definition of the set Λ and the definition of ϱ_n in (2.9) imply that $|\lambda|^2 \leq \varrho_n$ for $\lambda \in \Lambda$. Therefore for any $\lambda \in \Lambda$

$$\begin{aligned} \sum_{j=1}^n \lambda^2(j) \varsigma_{j,n} &\geq |\lambda|^2 \widehat{\varsigma}_n - \sigma_* \bar{\varrho}_{2,n} - |\lambda|^2 |\widehat{\varsigma}_n - \varsigma_n| \\ &\geq |\lambda|^2 \widehat{\varsigma}_n - \sigma_* \bar{\varrho}_{2,n} - \varrho_n |\widehat{\varsigma}_n - \varsigma_n|. \end{aligned}$$

Hence the desired inequality. \square

A.4 Properties of trigonometric basis

Lemma A.2. *For any $m \geq 0$,*

$$\sup_{N \geq 2} \sup_{x \in [0,1]} N^{-m} \left| \sum_{l=2}^N l^m \bar{\phi}_l(x) \right| \leq 2^m. \quad (\text{A.3})$$

Proof. Due to the properties of the trigonometric functions, we get

$$\begin{aligned} \sum_{l=2}^N l^m \bar{\phi}_l(x) &= \sum_{1 \leq l \leq N/2} (2l)^m \cos(4\pi l x) \\ &\quad - \sum_{1 \leq l \leq (N-1)/2} (2l+1)^m \cos(4\pi l x). \end{aligned}$$

This yields

$$\begin{aligned} \left| \sum_{l=2}^N l^m \bar{\phi}_l(x) \right| &\leq \left| \sum_{1 \leq l \leq (N-1)/2} ((2l+1)^m - (2l)^m) \cos(4\pi l x) \right| + N^m \\ &\leq \sum_{1 \leq l \leq (N-1)/2} ((2l+1)^m - (2l)^m) + N^m \\ &= \sum_{1 \leq l \leq (N-1)/2} \sum_{j=0}^{m-1} \binom{m}{j} (2l)^j + N^m. \end{aligned}$$

This implies (A.3).

□

Lemma A.3. For any function $S \in W_r^k$,

$$\sup_{n \geq 1} \sup_{1 \leq m \leq n-1} m^{2k} \left(\sum_{j=m+1}^n \theta_{j,n}^2 \right) \leq \frac{4r}{\pi^{2(k-1)}}. \quad (\text{A.4})$$

Proof. First, note that any function S from W_r^k can be represented by its Fourier series, i.e. $S = \sum_{j=1}^{\infty} \theta_j \phi_j$ with the coefficients defined by (3.4). By denoting the residual term for S as

$$\Delta_m(x) = S - \sum_{j=1}^m \theta_j \phi_j = \sum_{j=m+1}^{\infty} \theta_j \phi_j(x),$$

we obtain that

$$\sum_{j=m+1}^n \theta_{j,n}^2 = \inf_{\alpha_1, \dots, \alpha_m} \|S - \sum_{j=1}^m \alpha_j \phi_j\|_n^2 \leq \|\Delta_m\|_n^2.$$

Moreover, it is easy to deduce that

$$\begin{aligned}\|\Delta_m\|_n^2 &= n^{-1} \sum_{k=1}^n \Delta_m^2(x_k) = \sum_{k=1}^n \int_{x_{k-1}}^{x_k} \Delta_m^2(x_k) dx \\ &\leq 2 \int_0^1 \Delta_m^2(x) dx + 2 \sum_{k=1}^n \int_{x_{k-1}}^{x_k} (\Delta_m(x_k) - \Delta_m(x))^2 dx.\end{aligned}$$

The last term in this inequality we estimate as

$$\begin{aligned}(\Delta_m(x_k) - \Delta_m(x))^2 &= \left(\int_x^{x_k} \dot{\Delta}_m(z) dz \right)^2 \\ &\leq n^{-1} \int_{x_{k-1}}^{x_k} (\dot{\Delta}_m(z))^2 dz.\end{aligned}$$

Therefore,

$$\begin{aligned}\|\Delta_m\|_n^2 &\leq 2\|\Delta_m\|^2 + \frac{2}{n^2} \|\dot{\Delta}_m\|^2 \\ &= 2 \sum_{j=m+1}^{\infty} \theta_j^2 + \frac{2}{n^2} \sum_{j=m+1}^{\infty} \theta_j^2 \|\dot{\phi}_j\|^2.\end{aligned}$$

Now note that by the representation of the set W_r^k in the form (3.3) we can estimate the first term in the last inequality as

$$\sum_{j=m+1}^{\infty} \theta_j^2 = \sum_{j=m+1}^{\infty} \theta_j^2 \frac{a_j}{a_j} \leq \frac{r}{a_{m+1}} \leq \frac{r}{(\pi m)^{2k}}.$$

Similarly, we find that

$$\sum_{j=m+1}^{\infty} \theta_j^2 \|\dot{\phi}_j\|^2 \leq \sup_{j \geq m+1} \frac{\|\dot{\phi}_j\|^2}{a_j} r \leq \sup_{j \geq m+1} \frac{\|\dot{\phi}_j\|^2}{\|\phi_j^{(k)}\|^2} r \leq \frac{r}{(\pi m)^{2(k-1)}}.$$

Therefore, for $m \leq n$ we get that

$$\frac{1}{n^2} \sum_{j=m+1}^{\infty} \theta_j^2 \|\dot{\phi}_j\|^2 \leq \frac{r}{\pi^{2(k-1)} m^{2k}}.$$

This implies (A.4). \square

Lemma A.4. *Let $\xi_{j,n}$ be defined in (2.7) for the model (1.1). Then, for any real numbers v_1, \dots, v_n ,*

$$\mathbf{E} \left(\sum_{j=1}^n v_j \xi_{j,n} \right)^2 \leq \sigma_* \sum_{j=1}^n v_j^2. \quad (\text{A.5})$$

Proof. Due to the definition of $\xi_{j,n}$, one has

$$\sum_{j=1}^n v_j \xi_{j,n} = \sum_{l=1}^n \sigma_l \tilde{v}_l \xi_l$$

with

$$\tilde{v}_l = \frac{1}{\sqrt{n}} \sum_{j=1}^n v_j \phi_j(x_l).$$

Moreover

$$\begin{aligned} \mathbf{E} \left(\sum_{j=1}^n v_j \xi_{j,n} \right)^2 &= \sum_{l=1}^n \sigma_l^2(S) \tilde{v}_l^2 \leq \sigma_* \sum_{l=1}^n \tilde{v}_l^2 \\ &= \sigma_* \sum_{i,j=1}^n v_i v_j (\phi_i, \phi_j)_n. \end{aligned}$$

The orthogonality of the basis (ϕ_j) implies inequality (A.5). Hence Lemma A.4.

□

References

- [1] Akritas, M.G. and Van Keilegom, I. (2001) ANCOVA methods for heteroscedastic nonparametric regression models. *J. Amer. Statist. Assoc.* **96** 220-232.
- [2] Barron, A., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Related Fields.* **113** 301-413.
- [3] Cai, T. and Wang, L. (2008) Adaptive variance function estimation in heteroscedastic nonparametric regression. *Annals of Statistics*, 1-38 accepted.

- [4] Fourdrinier, D. and Pergamenshchikov, S. (2007) Improved selection model method for the regression with dependent noise. *Annals of the Institute of Statistical Mathematics*, **59** (3), p. 435-464.
- [5] Galtchouk, L. and Pergamenshchikov, S. (2004) Nonparametric sequential estimation of the drift in diffusion processes. *Mathematical Methods of Statistics*, **13**, 1, 25-49.
- [6] Galtchouk, L. and Pergamenshchikov, S. (2005) Efficient adaptive nonparametric estimation in heteroscedastic regression models. Preprint of the Strasbourg Louis Pasteur University, IRMA, 2005/020 available at <http://www.univ-rouen.fr/LMRS/Persopage/Pergamenchchikov>
- [7] Galtchouk, L. and Pergamenshchikov, S. (2006) Asymptotic efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Process*. **9**, 1-16.
- [8] Galtchouk, L. and Pergamenshchikov, S. (2007a) Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. Part 1. Sharp non-asymptotic oracle inequalities. *Mathematical Methods of Statistics*, submitted.
- [9] Galtchouk, L. and Pergamenshchikov, S. (2007b) Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. Part 2: Asymptotic efficiency. *Mathematical Methods of Statistics*, submitted.
- [10] Galtchouk, L. and Pergamenshchikov, S. (2007c) Adaptive nonparametric estimation in heteroscedastic regression models. Asymptotic efficiency. *Probab. Theory Related Fields*, submitted.
- [11] Goldfeld, S.M. and Quandt, R.E. (1972) *Nonlinear Methods in Econometrics*. North-Holland, London.
- [12] Kalifa, J. and Mallat, S. (2003) Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.* **31** 58-109.
- [13] Kneip, A. (1994) Ordered linear smoothers. *Ann. Statist.* **22** 835-866.
- [14] Massart, P. (2004) A non-asymptotic theory for model selection. *4ECM Stockholm 2004*. 309-323.

- [15] Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.* **13** 984-997.
- [16] Rohde, A. (2004) On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statistics & Decisions* **22** 235-243.
- [17] Young, S. G. and Bowman, A. W. (1995) Non-parametric Analysis of Covariance. *Biometrics* **51** 920-931.